PRTN #246

F. Tobagi
UCLA
January 13, 1978

On the Performance Analysis of Multihop Packet Radio Systems: Part I - The Design Problem

I. Introduction

The need for the sharing of computer resources by organizing these resources into computer networks has long been recognized and the feasibility of constructing such networks has been demonstrated by many successfully operating network systems. This economic sharing of computer resources has been made possible by the development of the packet switching technique [ROBE 70, KAHN 72, ABRA 73] whereby packet switches are interconnected by point-to-point data circuits according to some topological design. But when the number of communicating elements is sufficiently large and the overall traffic flow is small, the use of "packet broadcasting" techniques for interconnection becomes attractive, in that it simplifies the topological design considerably and provides very economic solutions. Moreover, economic studies [RODE 74] have clearly shown that, for geographically distributed networks, a significant part of the overall system cost is incurred by the local collection of data from, or dissemination of data to a large population of users. Today, with the proliferation of computer applications, computer resources have to be brought increasingly close to the individual; this makes it extremely desirable to create more economic techniques to also bring the communication capabilities closer to the individual. Again, radio offers an attractive solution; numerous papers have already appeared in the literature which discuss the advantages of using these techniques over wire communication

[KLEI 75A, KLEI 75B]. The ALOHA system, at the University of Hawaii [ABRA 70] appears to have been the first computer communication system utilizing radio and is an excellent illustration of the feasibility of the technique. In line with the objectives set forth in the design of the ALOHA system, and to allow the support of many applications which require several added features not existing in the ALOHA system, (such features as direct communication by a ground radio network between users over a wide geographical area, coexistence with possibly different systems in the same frequency band, anti-jam protection, etc...) the Advanced Research Projects Agency of the Department of Defense undertook a new effort, the packet radio system, in order to demonstrate the applicability of the packet radio concept in organizing computer resources into a computer communications network. The target system requirements have been well assessed by R. Kahn [KAHN 75]: "The system consists of terminals and stations linked together by line-of-sight radio repeaters. The stations are minicomputers which provide system control; the terminals are hand-held devices, I/O consoles, computers, sensors, etc. The repeaters are simple relay devices which provide network area coverage for terminals and for one or more stations... The system should be capable of meeting the requirements of mobile communication with computers including real-time speech communication... The system should be able to share a common frequency band with other (possibly different) systems ...The packet radio technology should be able to serve mobile users whether on land, at sea or in the air... The individual elements of the system should be constructed so that they can be installed in the field with little delay...The packet radio network should not require the presence of any personnel for its normal operation...The system should provide essentially error-free performance for computer communications..."

In this series of notes, we concern ourselves with the design and performance evaluation of packet radio networks. For this, we start by defining, in this first part, the design problem in its generality, identifying all key system variables. It will be apparent from the discussion that the general design problem is a very complex and intractable one; this suggests that a viable approach consists of considering and analyzing specific configurations which are intuitively appealing. The results obtained will constitute a valuable feedback by means of which the design defficiencies are detected and subsequently corrected, and new configurations invented.

II. The Design Problem

The design of a packet radio system meeting a given set of target requirements and satisfying given performance and reliability constraints is a very complex task. At first a feasibility study is required which either demonstrates that the requirements can be met by the present technology or dictates that research and improvements need to be made in various specific areas; in either case, there can be a number of alternatives to choose among; this leads to a second design phase, an optimization phase, which consists of selecting the best combination of choices. In its most general form, the design problem consists of minimizing the system cost over the many system variables discussed thereafter.

A. Network Topology

In landbased networks, node locations are given and the topological design problem consists of determining the best configuration connecting those nodes; in packet radio systems, however, the question is to determine

the number of devices (repeaters and stations), their locations, and their interconnections, needed to establish the required communications among sources and destinations. These variables are selected according to several factors: reliability, system load, frequency management... For example, the number of stations within the network is determined at first approximation by the amount of traffic to be handled; on the other hand, for reliability reasons the networks need to contain two or more stations; the allocation of responsibilities among the stations can be done either (i) by having one station assume the active role of controller while the other stations are kept in readiness for a backup mode, or (ii) by dynamically partitioning the control task mong them guaranteeing the backup of each station by the others in case of failure [KAHN 75]. The number of repeaters, on the other hand, is basically a function of: (i) the range of each repeater, essentially determined by its geographic setting and its effective radiated power, (ii) the repeater density required in achieving proper area coverage for communication with mobile hand-held terminals, and (iv) the frequency management and the communication mode adopted (point-to-point, broadcast, mixed...). The latter determines the interference pattern among the devices (it is discussed in more detail in the following paragraph).

B. Bandwidth Management

In digital nets using common carrier facilities, the system designer is given his choice of bit-rates and not of bandwidth. The common-carrier must worry about using his resources efficiently to provide a digital

channel. It is not so with the packet radio network. The r.f. spectrum is a limited resource and the system designer has to guarantee efficient use of the bandwidth.

The bandwidth required for a packet radio network will largely depend on the channel configuration adopted. If the allocated bandwidth is divided into a number of low speed channels, dedicated to pairs of devices with directional antennas in such a way that the r.f. signal interference is completely avoided, then we would simply be creating a point-to-point network! Such a configuration can be appropriate if one is in the presence of a known and non-varying traffic pattern; a proper design allocation will guarantee an efficient utilization of all channels at all times. An important drawback of this approach, however, is that it requires careful alignment and tailoring of the devices; this provides too rigid a configuration when the traffic load is varying or should one require that the system be extended to handle larger geographical areas without a major redesign effort.

The most advantageous property one can gain in using radio as a communication media is perhaps the possibility of broadcast transmission. When devices are equipped with omnidirectional antennas, one can get rapid and convenient deployment as well as the desirable area coverage for mobile terminals.

Computer communication is often characterized by a high ratio of peak to average traffic rates. When the traffic handled by the devices is so characterized, then providing channels on a user-pair basis proves to be extremely wasteful. The provision of a single high-speed channel to be shared by a large number of users provides a better utilization of

the bandwidth [TOBA 76]; it allows to take advantage of the "large number laws" whereby the demand at any instant can be approximated by the sum of the average demands of all users. With shared channels, however, we face the important problem of controlling the access to the channel by a geographically distributed set of message sources. The discussion of this issue is postponed until the next paragraph.

Traffic characteristics and other communications requirements may in some situations call for a network design with mixed modes. The bandwidth is partitioned into a number of channels some of which are shared while others are dedicated. The allocation of channels can be functional (channels used solely for a particular network function such as error control, monitoring...) and/or geographical to support the traffic requirements. In the latter case, the allocation has to be dynamic so that the statistical variations in traffic are accounted for. An illustration of a mixed mode network design is given by the "DAMN-FINE" concept introduced by Fralick [FRAL 73]. In this concept, the bandwidth is divided into a number of channels; each channel is time-shared among several mobile users and a repeater. Channel allocation is performed by an adaptive search and selection algorithm existing in each repeater and terminal, thus distributing the control and providing for an efficient dynamic allocation to match changing traffic and interference pattern.

C. Channel Access Policy

In conjunction with the proper selection of a channel configuration, the channel access procedure to be in use will have a great impact on the overall performance of the network.

With fixed bandwidth assignment, the time-bandwidth space is par-

titioned into slots which are allocated to the users in a static predetermined fashion. Fixed assignment takes two common forms: orthogonal, such as frequency division multiple access (FDMA) and synchronous time-division multiple access (TDMA), and "quasi-orthogonal", such as code division multiple access (CDMA). With fixed assignment the problem of accessing the channel is trivial. It is not so if the design calls for the usage of shared channels. Various methods are available. Some (such as contention and polling) require the presence of a central station performing the control. In a contention network, the terminal makes a request to transmit: if the channel is free, transmission goes ahead; if it is not free, the terminal must wait. The station schedules the transmissions either in a prearranged sequence (according to some scheduling scheme) or in the sequence in which the requests were made. In the polling technique, the station asks (polls) the terminals one by one as to whether they have anything to transmit. For this, the station may have a polling list giving the order in which terminals are polled. When a polling message is sent to the next terminal in sequence, and if the terminal has some data to transmit, it goes ahead; if not, a negative reply (or absence of reply) is received by the station and the next terminal is polled.

In the absence of central controllers, dynamic sharing can be accomplished through the so-called random access techniques. In this mode of sharing, however, one must be prepared to resolve conflicts which arise when more than one demand is simultaneously placed upon the channel; this is also referred to as interference. A simple scheme known as "pure ALOHA" [ABRA 70] permits users to transmit their packets anytime they desire

If within some appropriate time-out period they receive an acknowledgment from the destination, then they know that no conflicts occurred. Otherwise, they assume a collision occurred and they must retransmit. (To avoid continuously repeated conflicts, some scheme must be devised for introducing a random retransmission delay, spreading the conflicting packets over time; this parameter has been shown to have a great impact on the performance of the access mode, and its proper tuning is most crucial [KLEI 75B]. A second method for using the radio channel is to modify the completely unsynchronized use of the ALOHA channel by "slotting" time into segments whose duration is exactly equal to the transmission time of a single packet. If we require each user to start the transmission of his packets only at the beginning of a slot, then when two packets conflict, they will overlap completely rather than partially, providing an increase in channel efficiency. This method is referred to as "slotted ALOHA" [ROBE 72, ABRA 73, KLEI 73]. A third scheme, the "carrier sense multiple access" mode (CSMA) consists of reducing the level of interference (caused by overlapping packets) by allowing users to sense the carrier due to other users' transmissions; based on the information gained in this way about the state of the channel (busy or idle), the terminal takes an action prescribed by the particular CSMA protocol being used; (in particular, a terminal never transmits when it senses the channel busy). In [TOBA 74, KLEI 75B] two protocols referred to as nonpersistent and p-persistent CSMA were analyzed.

The selection of the proper channel access policy will largely depend on the behavior of these modes in the environment in question. It was shown [TOBA 76], for example, that if we are in the presence of a large

population of bursty users contending on the same channel then random access exhibits a performance far superior to fixed assignment or polling. However, centrally controlled techniques become more attractive when the contending population consists of a small number (≤ 5) of devices with buffering capabilities.

Split-channel reservation multiple access [TOBA 74, TOBA 76] is yet another channel access technique, suitable for packet radio environments, which calls upon both random access and central control. The available channel bandwidth is divided into two parts: one used to transmit control information (such as requests), the second used for the messages themselves. The request channel is operated in a random access mode; the requests are handled by a central controller which controls the access to the message channel. This technique represents an interesting scheme in that it is simple and efficient over a large range of system parameters The performance of the above access schemes has been compared in single-hop environments in which all devices are within range and in line-of-sight. Their performance in multi-hop environments is yet to be determined.

D. Modulation schemes

Just as with channel access, the modulation technique used is an important element in the design of a packet radio network. It has an important impact on network performance as well as on the utilization of the r.f. spectrum. Indeed, for a given required probability of bit error and a given transmitted power, modulation schemes determine the tradeoff relation between the range of a packet radio device (which affects the network topology) and the bit rate achievable (which affects the network throughput). While the simplest schemes achieve a relatively

inefficient use of power (which also constitutes a limiting resource when the equipments have to satisfy size and weight constraints), the most efficient ones require more complex receivers! From these points of view, the spread spectrum modulation scheme often represents a good compromise [FRAL 75]. In addition, it provides some kind of discrimination whereby one or more packets can be successfully detected in the presence of other packets; it also provides some degree of protection against unwanted interference and is desirable for jam-free operation. Finally, it renders coexistence with other systems feasible which may result in better utilization of the frequency spectrum. However, these advantages are obtained at the expense of a higher bandwidth requirement (possibly shared with other users). Discarding considerations of coexistence and vulnerability to unwanted interference (although in many design cases these may be the most critical factors), the designer is faced with a choice between spread-spectrum signalling and narrow-band (i.e., non-spread) systems: each presents different tradeoffs among some key system parameters such as the devices' radiated power, the r.f. bandwidth used, the channel data rate and the packet delay. A preliminary analysis made to that affect by Fralick [FRAL 74] has basically shown that the single-channel ALOHA spread-spectrum scheme is not an appealing choice in that the bandwidth required for given data-rate, throughput, and delay is much larger than that of the narrow-band system. The same result was also shown to be true for multi-channel spread-spectrum schemes with nonorthogonal ch nels.

E. Operational protocols: Routing policy, error control procedures and flow control

The principal functions of a packet communication system are to provide the transport of packets efficiently and reliably. Packets originate at traffic sources and have to be routed through the network to reach their destination. It is apparent that, in addition to all afore mentioned design factors, network performance will strongly be affected by the routing and flow control policies adopted.

If a point-to-point channel configuration is in use (dedicated channels with directional antennas), then the routing algorithms long devised for landbased store-and-forward networks will apply. They usually are classified as either deterministic or stochastic according to whether route computations are based upon a deterministic or stochastic decision rule [FULT 71]. A number of studies have investigated many of these routing algorithms, their performance and applicability. With shared broadcast channels, however, the nature of the problem becomes significantly different: packet transmissions are received by all nodes within range of the transmitting device; unless fixed predetermined routes are assigned and followed (a scheme which would defeat the purpose of broadcast transmission), the transmitted packet should carry, at each transmission, the next node's address. Equivalently, the routing algorithm can be defined as the decision at a receiving node as to whether to relay or ignore the packet. A number of routing schemes suitable for broadcast radio networks have been discussed by Gitman et. al. [GITM 76]. An undirected routing algorithm, also called broadcast routing, is based on an unconditional repeat of a correctly received packet as long as its handover number did not exceed a specified maximum. The maximum handover number determines the lifetime of a packet in the net. Looping is prevented by saving at each receiving repeater a unique

identifier of each packet for a specified length of time. This algorithm is inefficient in that it allows generation of a large number of duplicate packets and the delivery of the packet to its destination is guaranteed only to a certain extent. If the network structure is centralized*, then a directed routing algorithm can be used which utilizes the central station to periodically structure the network for efficient flow paths. A hierarchical point-to-point network can thus be formed by having the station assign to each repeater a label which defines its position in the hierarchy. The label carried in the packet header thus defines completely the path between the station and a repeater. Another class of routing is called the "directed broadcast" routing. Algorithms in this class are best suited to distributed radio networks in which communication has to be established between arbitrary source-destination pairs, and data need not flow through the station. To accomplish this task repeater nodes keep at all times routing information similar to that used in the ARPANET [KLEI 76]. The overhead incurred and the intelligence required by each node in performing the routing function may become significant.

The reliable transport of packets in a radio network can only be accomplished if proper error control procedures are in use. Indeed, in addition to random noise, errors in multi-access radio channels are due to multipath effects and interference caused by overlapping packets. Error detection code in conjunction with positive acknowledgment for each correct message is a reliable mechanism; however, the amount of overhead

---

*In a centralized network, all traffic passes through the station. Such a network architecture is best suitable when the radio network is used for local collection and distribution of traffic. The station then is the central computer facility or a gateway to other communications systems.

introduced and the degradation in network performance incurred still vary, here again, with the mode of operation. In [TOBA 78] for example, considering a single-hop system, we studied this degrading effect for two access schemes, slotted ALOHA and CSMA, and addressed ourselves to the problem of comparing two design alternatives: the common-channel configuration (a single channel for both information and error control traffic), and the split channel configurations. In particular, it was shown that the degradation in performance can be very significant when the acknowledgment packet size represents over 10% of the message packet!

In multi-hop networks we usually distinguish two types of acknowledgments: the hop-by-hop acknowledgment (HBH) and the end-to-end acknowledgment (ETE). HBH acknowledgments are issued by repeaters and guarantee the transport of a packet over a hop. They help decrease the delay incurred by end-to-end transmissions, but do so at the expense of added overhead. HBH acknowledging can be handled in several ways; HBH acknowledgments can be (i) piggy-backed (acknowledgment information is transmitted in the control portion of another packet destined to the sending repeater), (ii) passive (in that the relaying of a packet over a hop constitutes the acknowledgment for the transmission over the previous hop; this "echo-acknowledgment" mode is due to the omni-directional broadcast property), or (iii) active (whereby an acknowledgment packet is actually created and transmitted).

ETE acknowledging, in addition to guaranteeing a reliable communication between the end devices and the communicating processes, serves as a flow regulation mechanism; it has a great impact on network performance by preventing any source from overloading the network and causing serious congestion. The mechanism used also affects significantly

end-to-end message delays and buffer utilization at end devices. It is common to see several mechanisms in use in the same network depending on the particular application being supported.

In real systems, the proper functioning of operational protocols is dictated by their ability to adapt to changing system states, which in turn calls for the existence of efficient network monitoring functions at the various levels of protocols. Dynamic routing procedures, for example, require a thorough knowledge and a continuous update of the network connectivity and of the traffic load distribution in order to effect the appropriate route changes. Such monitoring functions can be distributed whereby devices can collect data regarding their own locality. In centralized networks, the station can be given the capability of centrally collecting data regarding the entire network and thus can perform both local and global decisions; it can be made responsible for route assignments and route changes, as well as the proper value assignment to all of the parameters throughout the network.

D. Repeater Design

The repeater is perhaps the most important element of a packet radio network. Almost all traffic flows through a number of repeaters on its way to destination. In addition to selecting the proper hardware to perform all r.f. functions (transmission, reception, error detection, bit synchronization, etc...) and providing the proper implementation of all previously discussed operational protocols, there is a number of design factors which are believed to have a crucial effect on the overall network performance. They are: the repeater's transmit power, its processing speed, and the storage capacity and its management. When the applicatio

require the devices to be rugged, light weighted, portable and operational for a period of time without attendance, then a number of constraints are placed on these parameters. The storage capacity has to be minimal and the power consumption has to be conservative; the latter affects both the processing speed (which in turn impacts the achievable throughput) and the radiated power (which in turn affects the range of the device and therefore the network topology). On the other hand, for a given topology, repeater's range affects network connectivity and the interference pattern...

To summarize, the design problem in radio networks can be expressed as follows:

Minimize the network cost subject to given constraints on throughput, delay, and reliability over the system variables:

1) Network topology: # of devices and their geographical setting
2) The bandwidth management: dedicated channels, shared channels, mixed modes
3) The channel access procedure: fixed assignment, centrally controlled schemes, random access
4) The modulation scheme: spread-spectrum, narrow band...
5) The operational protocols: routing policy, error control procedures, flow control and monitoring functions
6) Nodal design: storage capacity, buffer management, power requirement, processing speed

It is all apparent from the above discussion that the design of a PR network involves a large number of variables which interact in a very complex fashion. In its present form the solution is extremely hard to come by. A

reasonable approach to follow consists of first selecting those system variables which are obviously determined by some of the constraints. For example, for rapid deployment and easy communication among mobile devices, the entire system will employ omnidirectional antennas and will share a single high-speed channel via some random access scheme. Hopefully, this first step decreases significantly the variable space. The next step will consist of considering various specific configurations which are intuitively appealing; these configurations are then analyzed in order to

(i) identify the key parameters which affect the performance

(ii) determine the conditions under which the a priori specified constraints can be satisfied

This step represents an iterative process in that the results obtained from the analysis of some network configuration constitutes a valuable feedback process by the means of which the design deficiencies are detected and subsequently corrected, and new configurations are invented.

In the forthcoming notes, we shall execute this second step by focusing on the analysis of specific network configurations (star networks, fully connected networks) operating under the slotted ALOHA and the nonpersistent CSMA modes.

References

1 ABRA 70 Abramson, N. "Another Alternative for Computer Communications," AFIPS Conference Proceedings, FJCC 70, pp. 695-702.

2 ABRA 73A Abramson, N. and F. Kuo (co-editors), Computer Communication Networks. New York: Prentice-Hall Pub. Co., 1973

3 ABRA 73B Abramson, N., "Packet Switching with Satellites," in 1973 Nat. Comput. Conf., AFIPS Conference Proceedings, vol. 42, New York, 1973, pp. 695-702.

4 FRAL 74 Fralick, S. and J. K. Leung, "Study of Throughput and Delay of Spread-Spectrum Multiple Access Modes," Packet Radio Note 2, SRI International, Menlo Park, California, April 1974.

5 FRAL 75 Fralick, S. and J. Garrett, "A Technology for Packet Radio," AFIPS Conference Proceedings, vol. 44, 1975, AFIPS Press, Montvale, N.J.

6 FRAL 73 Fralick, S., "The DAMN (-FINE) Concept," Packet Radio Temporary Note #82, SRI International, Menlo Park, California, Nov. 1973.

7 FULT 71 Fultz, G. and L. Kleinrock, "Adaptive Routing Techniques for Store-and-Forward Computer Communications Networks," Proc. of the Internat. Conf. on Commun., pp. 39-1 to 39-8, Montreal, Canada, 1971.

8 GITM 76 Gitman, I., R. M. Van Slyke and H. Frank, "Routing in Packet-Switching Broadcast Radio Networks," IEEE Trans. Commun., vol. COM-24, Aug. 1976.

9 KAHN 72 Kahn, R. E., "Resource Sharing Computer Communication Networks," IEEE Proceedings, Nov. 1972, pp. 1397-1407.

10 KAHN 75 Kahn, R. E., "The Organization of Computer Resources into a Packet Radio Network," AFIPS Conference Proceedings, NCC 75, pp. 177-186.

11 KLEI 73 Kleinrock, L. and S. Lam, "Packet Switching in a Slotted Satellite Channel," in Nat. Comput. Conf., AFIPS Conference Proceedings, vol. 42, Montvale, N.J.: AFIPS Press, 1973, pp. 703-710.

12 KLEI 75A Kleinrock, L. and S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," IEEE Trans. Commun., vol. COM-23, pp. 410-423, Apr. 1975.

13 KLEI 75B Kleinrock, L. and F. Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple Access Modes and Their Throughput-Delay Characteristics," IEEE Trans. Commun., vol. COM-23, pp. 1400-1416, Dec. 1975.

14 KLEI 76 Kleinrock, L., Queueing Systems, vol. II, Computer Applications New York: Wiley-Interscience, 1976.

15 ROBE 70 Roberts, L. G. and B. D. Wessler, "Computer Network Development to Achieve Resource Sharing," AFIPS Conference Proceedings, SJCC 70, pp. 543-549.

16 ROBE 72 Roberts, L. G., "ALOHA Packet System with and without Slots and Capture," ARPA Network Inform. Center, Stanford Res. Inst., Menlo Park, Calif., ASS Note 8 (NIC 11290), June 1972.

17 ROBE 74 Roberts, L. "Data by the Packet," IEEE Spectrum, February 1974.

18 TOBA 74 Tobagi, F., "Random Access Techniques for Data Transmission over Packet Switched Radio Networks," Ph.D. dissertation, Comput. Sci. Dept., School of Engr. and Appl. Sci., Univ., of California, Los Angeles, rep. UCLA-ENG 7499, Dec. 1974.

19 TOBA 76 Tobagi, F. and L. Kleinrock, "Packet Switching in Radio Channels: Part III - Polling and (Dynamic) Split-Channel Reservation Multiple Access," IEEE Trans. Commun., vol. COM-24, pp. 832-845, Aug. 1976.

20 TOBA 78 Tobagi, F. and L. Kleinrock, "The Effect of Acknowledgment Traffic on the Capacity of Packet-Switched Radio Channels," to appear in the IEEE Trans. Commun.